

# Max-Weight Achieves the Exact $[O(1/V), O(V)]$ Utility-Delay Tradeoff Under Markov Dynamics

Longbo Huang, Michael J. Neely

**Abstract**—In this paper, we show that the Quadratic Lyapunov function based Algorithm (QLA, also known as MaxWeight or Backpressure) achieves an exact  $[O(1/V), O(V)]$  utility-delay tradeoff in stochastic network optimization problems with Markovian network dynamics. Note that though the QLA algorithm has been extensively studied, most of the performance results are obtained under i.i.d. network randomness, and it has not been formally proven that QLA achieves the exact  $[O(1/V), O(V)]$  utility-delay tradeoff under Markov dynamics. Our analysis uses a combination of duality theory and a variable multi-slot Lyapunov drift argument. The variable multi-slot Lyapunov drift argument here is different from previous multi-slot drift analysis, in that the slot number is a random variable corresponding to the renewal time of the network randomness. This variable multi-slot drift argument not only allows us to obtain an exact  $[O(1/V), O(V)]$  tradeoff, but also allows us to state the performance of QLA in terms of explicit parameters of the network dynamic process.

**Index Terms**—Queueing, Dynamic Control, Lyapunov analysis, Stochastic Optimization

## I. INTRODUCTION

In this paper, we show that the Quadratic Lyapunov function based Algorithm (QLA, also known as the MaxWeight algorithm) [1] achieves an exact  $[O(1/V), O(V)]$  utility-delay tradeoff in the following general stochastic network optimization problem. We are given a discrete time stochastic network. The network state, which describes the network randomness, such as the network channel condition or the random arrivals, is time varying according to some Markov process. A network controller performs some action based on the observed network state at every time slot. The chosen action incurs a cost,<sup>1</sup> but also serves some amount of traffic and possibly generates new traffic for the network. This traffic causes congestion, and thus leads to backlogs at nodes in the network. The goal of the controller is to minimize its time average cost subject to the constraint that the time average total backlog in the network is finite.

This is a very general framework and includes a wide class of networking problems, ranging from flow utility maximization [2], energy minimization [3], network pricing [4] to cognitive radio applications [5] etc. Also, many

techniques have also been applied to this problem (see [6] for a survey). Among the many techniques that have been adopted, the family of Quadratic Lyapunov function based Algorithms (QLA) [1] are recently receiving much attention, due to their provable performance guarantees, robustness to stochastic network conditions, and most importantly, their ability to achieve the desired performance *without requiring any statistical knowledge* of the underlying randomness in the network. When the network state is i.i.d., it has been proven in [1] that QLA can achieve a utility that is within  $O(1/V)$  of the optimal utility for any  $V \geq 1$  for general network optimization problems, while guaranteeing an  $O(V)$  network delay. Two works [7] [8] construct algorithms to achieve an  $[O(1/V), O(\log(V))]$  utility-delay tradeoff using exponential Lyapunov functions. The recent work [9] also develops the Fast-QLA (FQLA) algorithm based on quadratic Lyapunov functions to achieve an  $[O(1/V), O((\log(V))^2)]$  tradeoff.

When the network state is Markovian, it has been shown that when the network backlogs are deterministically bounded, QLA can also achieve utilities within  $O(\log(V)/V)$  to the optimal values [4] [5], while guaranteeing that the average delay is  $O(V)$ . Without such deterministic queueing bounds, it has recently been shown that QLA achieves an  $[O(\epsilon + \frac{T_\epsilon}{V}), O(V)]$  tradeoff under Markovian network states [10], where  $\epsilon > 0$  and  $T_\epsilon$  represent the proximity to the optimal value and the “convergence time” of the QLA algorithm to that proximity, respectively. However, there has not been any proof showing that QLA achieves the exact  $[O(1/V), O(V)]$  utility-delay tradeoff under Markovian network dynamics.

In this paper, we present the first proof of the exact  $[O(1/V), O(V)]$  tradeoff of the QLA algorithm under Markovian network dynamics. To establish the result, we use a combination of duality theory and a variable multi-slot Lyapunov drift argument. Different from previous multi-slot drift arguments, e.g., [1], where the drift is usually computed over a fixed number of slots, the slot number here is a random variable corresponding to the renewal time of the network dynamic process. This  $[O(1/V), O(V)]$  tradeoff result contributes to a better understanding of the QLA algorithm performance and enables more precise resource allocation decisions in network optimization problems. The result can also be combined with the recent result developed in [9] to show that the FQLA algorithm achieves an  $[O(1/V), O((\log(V))^2)]$  tradeoff for stochastic network optimization problems with Markovian network dynamics, and is thus the first known algorithm that can ensure a poly-logarithmic delay performance when pushing the utility performance to within  $O(1/V)$  of the optimal in this Markovian case.

Longbo Huang (web: <http://www-scf.usc.edu/~longbohu>) and Michael J. Neely (web: <http://www-rcf.usc.edu/~mjneely>) are with the Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90089, USA.

This material is supported in part by one or more of the following: the DARPA IT-MANET program grant W911NF-07-0028, the NSF grant OCE 0520324, the NSF Career grant CCF-0747525.

<sup>1</sup>Since cost minimization is mathematically equivalent to utility maximization, below we will use cost and utility interchangeably

This paper is organized as follows. In Section II, we set up our notations. We then present our system model in Section III. We review the QLA algorithm in Section IV. The performance results of QLA under the Markovian network dynamics are obtained in Section V.

## II. NOTATIONS

Here we specify our notations.  $\mathbb{R}$  represents the set of real numbers.  $\mathbb{R}_+$  (or  $\mathbb{R}_-$ ) represents the set of nonnegative (or non-positive) real numbers.  $\mathbb{R}^n$  (or  $\mathbb{R}_+^n$ ) represents the set of  $n$  dimensional *column* vectors, with each element being in  $\mathbb{R}$  (or  $\mathbb{R}_+$ ). **bold** symbols  $\mathbf{a}$  and  $\mathbf{a}^T$  represent *column* vector and its transpose.  $\mathbf{a} \succeq \mathbf{b}$  indicates that vector  $\mathbf{a}$  is entrywise no less than vector  $\mathbf{b}$ .  $\|\mathbf{a} - \mathbf{b}\|$  is the Euclidean distance of  $\mathbf{a}$  and  $\mathbf{b}$ .  $\mathbf{0}$  is the column vector with all elements being 0.

## III. SYSTEM MODEL

In this section, we specify the general network model we use. We consider a network controller that operates a network with the goal of minimizing the time average cost, subject to the queue stability constraint. The network is assumed to operate in slotted time, i.e.,  $t \in \{0, 1, 2, \dots\}$ . We assume there are  $r \geq 1$  queues in the network.

### A. Network State

In every slot  $t$ , we use  $S(t)$  to denote the current network state, which indicates the current network parameters, such as a vector of channel conditions for each link, or a collection of other relevant information about the current network channels and arrivals. We assume that  $S(t)$  evolves according to a general irreducible and aperiodic Markov chain with countably many states and denote its state space by  $\mathcal{S} = \{s_1, s_2, s_3, \dots\}$ . We assume  $S(t)$  has a well defined steady state distribution, and let  $\pi_{s_i}$  denote its steady state probability of being in state  $s_i$ . Note that in this case, by Theorem 3 in Chapter 5 of [11], the existence of a steady state distribution  $\pi$  implies that all the states are positive recurrent, hence  $\pi_{s_i} > 0$  for all  $i$ .

### B. The Cost, Traffic, and Service

At each time  $t$ , after observing  $S(t) = s_i$ , the controller chooses an action  $x(t)$  from a set  $\mathcal{X}^{(s_i)}$ , i.e.,  $x(t) = x^{(s_i)}$  for some  $x^{(s_i)} \in \mathcal{X}^{(s_i)}$ . The set  $\mathcal{X}^{(s_i)}$  is called the feasible action set for network state  $s_i$  and is assumed to be time-invariant and compact for all  $s_i \in \mathcal{S}$ . The cost, traffic, and service generated by the chosen action  $x(t) = x^{(s_i)}$  are as follows:

- The chosen action has an associated cost given by the cost function  $f(t) = f(s_i, x^{(s_i)}) : \mathcal{X}^{(s_i)} \mapsto \mathbb{R}_+$  (or  $\mathcal{X}^{(s_i)} \mapsto \mathbb{R}_-$  in reward maximization problems);
- The amount of traffic generated by the action to queue  $j$  is determined by the traffic function  $A_j(t) = A_j(s_i, x^{(s_i)}) : \mathcal{X}^{(s_i)} \mapsto \mathbb{R}_+$ , in units of packets;
- The amount of service allocated to queue  $j$  is given by the rate function  $\mu_j(t) = \mu_j(s_i, x^{(s_i)}) : \mathcal{X}^{(s_i)} \mapsto \mathbb{R}_+$ , in units of packets;

Note that  $A_j(t)$  includes both the exogenous arrivals from outside the network to queue  $j$ , and the endogenous arrivals from other queues, i.e., the transmitted packets from other queues,

to queue  $j$ . We assume the functions  $f(s_i, \cdot)$ ,  $\mu_j(s_i, \cdot)$  and  $A_j(s_i, \cdot)$  are time-invariant, their magnitudes are uniformly upper bounded by some constant  $\delta_{max} \in (0, \infty)$  for all  $s_i, j$ , and they are known to the network operator. We also assume that there exists a set of actions  $\{x_k^{(s_i)}\}_{k=1,2,\dots,r+2}$  with  $x_k^{(s_i)} \in \mathcal{X}^{(s_i)}$  for all  $s_i$ , and a set of variables  $\{\vartheta_k^{(s_i)}\}_{k=1,2,\dots,r+2}$  with  $\sum_k \vartheta_k^{(s_i)} = 1$  and  $\vartheta_k^{(s_i)} \geq 0$  for all  $s_i$  and  $k$  such that:

$$\sum_{s_i} \pi_{s_i} \left\{ \sum_k \vartheta_k^{(s_i)} [A_j(s_i, x_k^{(s_i)}) - \mu_j(s_i, x_k^{(s_i)})] \right\} \leq -\eta, \quad (1)$$

for some  $\eta > 0$  for all  $j$ . That is, the queue stability constraints are feasible with  $\eta$ -slackness. Thus, there exists a stationary randomized policy that stabilizes all queues (where  $\vartheta_k^{(s_i)}$  represents the probability of choosing action  $x_k^{(s_i)}$  when  $S(t) = s_i$ ). In the following, we use  $\mathbf{A}(t) = (A_1(t), A_2(t), \dots, A_r(t))^T$  and  $\boldsymbol{\mu}(t) = (\mu_1(t), \mu_2(t), \dots, \mu_r(t))^T$  to denote the arrival and service vectors at time  $t$ . It is easy to see from above that if we define:

$$B = \sqrt{r} \delta_{max}, \quad (2)$$

then  $\|\mathbf{A}(t) - \boldsymbol{\mu}(t)\| \leq B$  for all  $t$ .

### C. Queueing, Average Cost, and the Stochastic Problem

Let  $\mathbf{q}(t) = (q_1(t), \dots, q_r(t))^T \in \mathbb{R}_+^r$ ,  $t = 0, 1, 2, \dots$  be the queue backlog vector process of the network, in units of packets. We assume the following queueing dynamics:

$$q_j(t+1) = \max[q_j(t) - \mu_j(t), 0] + A_j(t) \quad \forall j, \quad (3)$$

and  $\mathbf{q}(0) = \mathbf{0}$ . By using (3), we assume that when a queue does not have enough packets to send, null packets are transmitted. In this paper, we adopt the following notion of queue stability:

$$\mathbb{E} \left\{ \sum_{j=1}^r q_j \right\} \triangleq \limsup_{t \rightarrow \infty} \frac{1}{t} \sum_{\tau=0}^{t-1} \sum_{j=1}^r \mathbb{E} \{ q_j(\tau) \} < \infty. \quad (4)$$

We also use  $f_{av}^\Pi$  to denote the time average cost induced by an action-choosing policy  $\Pi$ , defined as:

$$f_{av}^\Pi \triangleq \limsup_{t \rightarrow \infty} \frac{1}{t} \sum_{\tau=0}^{t-1} \mathbb{E} \{ f^\Pi(\tau) \}, \quad (5)$$

where  $f_{av}^\Pi(\tau)$  is the cost incurred at time  $\tau$  by policy  $\Pi$ . We call an action-choosing policy *feasible* if at every time slot  $t$ , it only chooses actions from the feasible action set  $\mathcal{X}^{(S(t))}$ . We then call a feasible action-choosing policy under which (4) holds a *stable* policy, and use  $f_{av}^*$  to denote the optimal time average cost over all stable policies. In every slot, the network controller observes the current network state  $S(t)$  and chooses a control action, with the goal of minimizing time average cost subject to network stability. This goal can be mathematically stated as: **(P1)**  $\min : f_{av}^\pi, \text{ s.t. (4)}$ . In the rest of the paper, we will refer to problem **(P1)** as *the stochastic problem*.

#### IV. QLA AND THE DETERMINISTIC PROBLEM

In this section, we first review the quadratic Lyapunov function based algorithms (the QLA algorithm) [1] for solving the stochastic problem. Then we define the *deterministic problem* and its dual problem. We then also discuss some properties of the dual function. The dual problem and the properties of the dual function will be used later for analyzing the performance of QLA.

We first recall the QLA algorithm [1] as follows.

**QLA:** Initialize the parameter  $V \geq 1$ . At every time slot  $t$ , observe the current network state  $S(t)$  and the backlog  $q(t)$ . If  $S(t) = s_i$ , choose  $x^{(s_i)} \in \mathcal{X}^{(s_i)}$  that solves the following:

$$\begin{aligned} \max : \quad & -Vf(s_i, x) + \sum_{j=1}^r q_j(t) [\mu_j(s_i, x) - A_j(s_i, x)] \\ \text{s.t.} \quad & x \in \mathcal{X}^{(s_i)}. \end{aligned} \quad (6)$$

Depending on the problem structure, (6) can usually be decomposed into separate parts that are easier to solve, e.g., [3], [4]. Also, when the network state process  $S(t)$  is i.i.d., it has been shown in [1] that,

$$f_{av}^{QLA} = f_{av}^* + O(1/V), \quad \bar{q}^{QLA} = O(V), \quad (7)$$

where  $f_{av}^{QLA}$  and  $\bar{q}^{QLA}$  are the expected average cost and the expected time average network backlog size under QLA, respectively. When  $S(t)$  is Markovian, it has been shown in, e.g., [4] and [5] that QLA achieves an  $[O(\log(V)/V), O(V)]$  utility-delay tradeoff if the queue sizes are deterministically upper bounded by  $\Theta(V)$  for all time. Without this deterministic backlog bound, it has recently been shown that QLA achieves an  $[O(\epsilon + \frac{T_\epsilon}{V}), O(V)]$  tradeoff under Markov  $S(t)$  processes, where  $\epsilon$  and  $T_\epsilon$  represent the proximity to the optimal value and the ‘‘convergence time’’ of the QLA algorithm for this proximity [10]. However, this latter tradeoff is less explicit, and it is common that when  $S(t)$  is Markovian,  $T_\epsilon = \Omega(\log(\frac{1}{\epsilon}))$ , in which case we again have an  $[O(\frac{\log(V)}{V}), O(V)]$  tradeoff when  $\epsilon = 1/V$ .

We also recall the *deterministic problem* defined in [9]:

$$\begin{aligned} \min : \quad & \mathcal{F}(\mathbf{x}) \triangleq V \sum_{s_i} \pi_{s_i} f(s_i, x^{(s_i)}) \\ \text{s.t.} \quad & \mathcal{A}_j(\mathbf{x}) \triangleq \sum_{s_i} \pi_{s_i} A_j(s_i, x^{(s_i)}) \\ & \leq \mathcal{B}_j(\mathbf{x}) \triangleq \sum_{s_i} \pi_{s_i} \mu_j(s_i, x^{(s_i)}) \quad \forall j \\ & x^{(s_i)} \in \mathcal{X}^{(s_i)} \quad \forall i = 1, 2, \dots \end{aligned} \quad (8)$$

where  $\pi_{s_i}$  corresponds to the steady state probability of  $S(t) = s_i$  and  $\mathbf{x} = (x^{(s_1)}, \dots, x^{(s_M)})^T$ . The dual problem of (8) can be obtained as follows:

$$\max : \quad g(\gamma), \quad \text{s.t.} \quad \gamma \succeq \mathbf{0}, \quad (9)$$

where  $g(\gamma)$  is called the dual function and is defined as:

$$\begin{aligned} g(\gamma) = \inf_{x^{(s_i)} \in \mathcal{X}^{(s_i)}} \sum_{s_i} \pi_{s_i} \left\{ Vf(s_i, x^{(s_i)}) \right. \\ \left. + \sum_j \gamma_j [A_j(s_i, x^{(s_i)}) - \mu_j(s_i, x^{(s_i)})] \right\}. \end{aligned} \quad (10)$$

Here  $\gamma = (\gamma_1, \dots, \gamma_r)^T$  is the *Lagrange multiplier* of (8). It is well known that  $g(\gamma)$  in (10) is concave in the vector  $\gamma$ , and hence the problem (9) can usually be solved efficiently, particularly when cost functions and rate functions are separable over different network components. It is also well known that in many situations, the optimal value of (9) is the same as the optimal value of (8) and in this case we say that there is no duality gap [12]. However, despite the fact that the problem (8) may be non-convex, in which case the duality gap is usually nonzero, our first result shows that the dual problem (9) gives the exact value of  $Vf_{av}^*$ , where  $f_{av}^*$  is the optimal time average cost for the stochastic problem. Below,  $\gamma_V^* = (\gamma_{V1}^*, \gamma_{V2}^*, \dots, \gamma_{Vr}^*)^T$  denotes an optimal solution of the dual problem (9) with the corresponding  $V$  parameter.

**Theorem 1:** Let  $\gamma_V^*$  be an optimal solution of the dual problem (9). We have:

$$g(\gamma_V^*) = Vf_{av}^*. \quad (11)$$

*Proof:* See Appendix A. ■

The following corollary is immediate and will be useful for our following analysis.

**Corollary 1:** For any  $\gamma \succeq \mathbf{0}$ , we have:

$$g(\gamma) \leq Vf_{av}^*. \quad (12)$$

In the following, we also define the functions  $g_{s_i}(\gamma)$  for each  $s_i = s_1, s_2, \dots$  as follows:

$$\begin{aligned} g_{s_i}(\gamma) = \inf_{x^{(s_i)} \in \mathcal{X}^{(s_i)}} \left\{ Vf(s_i, x^{(s_i)}) \right. \\ \left. + \sum_j \gamma_j [A_j(s_i, x^{(s_i)}) - \mu_j(s_i, x^{(s_i)})] \right\}. \end{aligned} \quad (13)$$

That is, the  $g_{s_i}(\cdot)$  function is the dual function of (8) when the network has only one single network state  $s_i$ , i.e., the network condition is deterministically described by  $s_i$ . It is easy to see from (10) and (13) that:

$$g(\gamma) = \sum_{s_i} \pi_{s_i} g_{s_i}(\gamma). \quad (14)$$

Also, the term  $\mathbf{G}_{\gamma}^{(s_i)} = (G_{\gamma,1}^{(s_i)}, G_{\gamma,2}^{(s_i)}, \dots, G_{\gamma,r}^{(s_i)})^T$  with:

$$G_{\gamma,j}^{(s_i)} = [-\mu_j(s_i, x_{\gamma}^{(s_i)}) + A_j(s_i, x_{\gamma}^{(s_i)})], \quad (15)$$

is called the subgradient of the  $g_{s_i}(\cdot)$  function at the point  $\gamma$  [12]. It is known that for any other  $\hat{\gamma} \in \mathbb{R}^r$ , we have:

$$(\hat{\gamma} - \gamma)^T \mathbf{G}_{\gamma}^{(s_i)} \geq g_{s_i}(\hat{\gamma}) - g_{s_i}(\gamma). \quad (16)$$

Using the fact that  $\|\mathbf{G}_{\gamma}^{(s_i)}\| \leq B$ , (16) also implies:

$$g_{s_i}(\hat{\gamma}) - g_{s_i}(\gamma) \leq B\|\hat{\gamma} - \gamma\|. \quad (17)$$

#### V. PERFORMANCE OF QLA UNDER MARKOVIAN DYNAMICS

In this section, we prove that under the Markovian network state dynamics, QLA achieves an exact  $[O(\frac{1}{V}), O(V)]$  utility-delay tradeoff for the stochastic problem. This is the first formal proof of this result. It generalizes the  $[O(\frac{1}{V}), O(V)]$  performance result of QLA in the i.i.d. case in [1]. To prove the result, we use a variable multi-slot Lyapunov drift argument.

Different from previous multi-slot drift arguments, e.g., [13] and [10], where the drift is usually computed over a fixed number of slots, the slot number here is a random variable corresponding to the return time of the network states. As we will see, this variable multi-slot drift analysis allows us to obtain the exact  $[O(\frac{1}{V}), O(V)]$  utility-delay tradeoff for QLA. Moreover, it also allows us to state QLA's performance in terms of explicit parameters of the Markovian  $S(t)$  process.

In the following, we define  $T_i(t_0)$  to be the first return time of  $S(t)$  to state  $s_i$  given that  $S(t_0) = s_i$ , i.e.,

$$T_i(t_0) = \inf\{T > 0, \text{ s.t. } S(t_0 + T) = s_i \mid S(t_0) = s_i\}.$$

We see that  $T_i(t_0)$  has the same distribution for all  $t_0$ . Thus, we will use  $\overline{T}_i$  to denote the expected value of  $T_i(t)$  for any  $t$  s.t.  $S(t) = s_i$  and use  $\overline{T}_i^2$  to denote its second moment. By Theorem 3 in Chapter 5 of [11], we have for all states  $s_i$  that:

$$\overline{T}_i = \frac{1}{\pi_{s_i}} < \infty, \quad (18)$$

i.e., the expected return time of any state  $s_i$  is finite. In the following, we also use  $T_{ji}(t_0)$  to denote the first hitting time for  $S(t)$  to enter the state  $s_i$  given that  $S(t_0) = s_j$ . It is again easy to see that  $T_{ji}(t_0)$  has the same distribution at all  $t_0$ . Hence we similarly use  $\overline{T}_{ji}$  and  $\overline{T}_{ji}^2$  to denote its first and second moments. Throughout the paper, we make the following assumption:

**Assumption 1:** There exists a state  $s_1$  such that:

$$\overline{T}_{j1}^2 < \infty, \quad \forall j.$$

That is, starting from any state  $s_j$  (including  $s_1$ ), the random time needed to get into state  $s_1$  has a finite second moment. This condition is not very restrictive and can be satisfied in many cases, e.g., when  $\mathcal{S}$  is finite.

We now have the following theorem summarizing QLA's performance under the Markovian network state dynamics:

**Theorem 2:** Suppose (1) holds. Then under the Markovian network state process  $S(t)$ , the QLA algorithm achieves the following:

$$f_{av}^{QLA} \leq f_{av}^* + \frac{CB^2}{V\overline{T}_1}, \quad (19)$$

$$\sum_{j=1}^r q_j \leq \frac{CB^2 + \overline{T}_1 V \delta_{max}}{\eta} + \frac{DB^2}{2}, \quad (20)$$

where  $\eta > 0$  is the slack parameter defined in (1) in Section III-B, and  $C, D$  are defined as:

$$C = \overline{T}_1^2 + \overline{T}_1, \quad D = \overline{T}_1^2 - \overline{T}_1, \quad (21)$$

i.e.,  $C$  and  $D$  are the sum and difference of the first and second moments of the return time associated with  $s_1$ .

Note that  $\eta, C, D = \Theta(1)$  in (20), i.e., independent of  $V$ . Hence Theorem 2 shows that QLA indeed achieves an exact  $[O(1/V), O(V)]$  utility-delay tradeoff for general stochastic network optimization problems with Markovian network dynamics. Although our bounds may be loose when the number of states is large, we note that Theorem 2 also applies to the case when  $S(t)$  evolves according to a Markov modulated i.i.d. process, in which case there is a Markov chain of only

a few states, but in each Markov state, there can be many i.i.d. randomness. For example, suppose  $S(t)$  is i.i.d. with  $10^4$  states. Then we can view  $S(t)$  as having one Markov state, but within the Markov state, it has  $10^4$  i.i.d. random choices. In this case, Theorem 2 will apply with  $C = 2$  and  $D = 0$ . These Markov modulated processes can easily be incorporated into our analysis by taking expectation over the i.i.d. randomness of the current Markov state in Equation (22). These Markov modulated processes are important in stochastic modeling and include the *ON/OFF* processes for modeling time-correlated arrivals processes, e.g., [14].

*Proof:* (Theorem 2) To prove the theorem, we first define the Lyapunov function  $L(t) = \frac{1}{2} \sum_{j=1}^r q_j^2(t)$ . By using the queueing dynamic equation (3), it is easy to obtain that:

$$\frac{1}{2} q_j^2(t+1) - \frac{1}{2} q_j^2(t) \leq \delta_{max}^2 + q_j(t)[A_j(t) - \mu_j(t)].$$

Summing over all  $j = 1, \dots, r$  and adding to both sides the term  $Vf(t)$ , we obtain:

$$\begin{aligned} L(t+1) - L(t) + Vf(t) \\ \leq B^2 + \left\{ Vf(t) + \sum_{j=1}^r q_j(t)[A_j(t) - \mu_j(t)] \right\}. \end{aligned} \quad (22)$$

We see from (6) then given the network state  $S(t)$ , QLA chooses an action to *minimize the right-hand side (RHS) at time  $t$* . Now compare the term in  $\{\}$  in the RHS of (22) with (13), we see that we indeed have:

$$L(t+1) - L(t) + Vf^Q(t) \leq B^2 + g_{S(t)}(\mathbf{q}(t)), \quad (23)$$

where we use  $f^Q(t) = f(x^{QLA}(t))$  to denote the utility incurred by QLA's action at time  $t$ , and  $g_{S(t)}(\cdot)$  is the function (13) with the network state being  $S(t)$ .

(Part A: Proof of Utility) We first prove the utility performance. Consider  $t = 0$  and first assume that  $S(0) = s_1$ . Summing up the inequality (23) from time  $t = 0$  to time  $t = T_1(0) - 1$ , we have:

$$\begin{aligned} L(T_1(0)) - L(0) + \sum_{t=0}^{T_1(0)-1} Vf^Q(t) \leq T_1(0)B^2 \\ + \sum_{t=0}^{T_1(0)-1} g_{S(t)}(\mathbf{q}(t)). \end{aligned}$$

This can be rewritten as:

$$\begin{aligned} L(T_1(0)) - L(0) + \sum_{t=0}^{T_1(0)-1} Vf^Q(t) \leq T_1(0)B^2 \\ + \sum_{t=0}^{T_1(0)-1} g_{S(t)}(\mathbf{q}(0)) + \sum_{t=0}^{T_1(0)-1} [g_{S(t)}(\mathbf{q}(t)) - g_{S(t)}(\mathbf{q}(0))]. \end{aligned} \quad (24)$$

Using (17) and the fact that  $\|\mathbf{q}(t+\tau) - \mathbf{q}(t)\| \leq \tau B$ , we see that the final term can be bounded by:

$$\begin{aligned} \left| \sum_{t=0}^{T_1(0)-1} [g_{S(t)}(\mathbf{q}(t)) - g_{S(t)}(\mathbf{q}(0))] \right| \leq \sum_{t=0}^{T_1(0)-1} tB^2 \\ = \left[ \frac{1}{2}(T_1(0))^2 - \frac{1}{2}T_1(0) \right] B^2. \end{aligned}$$



Plugging this into (24), and letting  $\hat{C} = \frac{1}{2}(T_1(0))^2 + \frac{1}{2}T_1(0)$ , we obtain:

$$\begin{aligned} L(T_1(0)) - L(0) &+ \sum_{t=0}^{T_1(0)-1} V f^Q(t) \\ &\leq \hat{C} B^2 + \sum_{t=0}^{T_1(0)-1} g_{S(t)}(\mathbf{q}(0)) \\ &= \hat{C} B^2 + \sum_{s_i} n_{s_i}^{T_1(0)}(0) g_{s_i}(\mathbf{q}(0)). \end{aligned} \quad (25)$$

Here  $n_{s_i}^{T_1(0)}(0)$  denotes the number of times the network state  $s_i$  appears in the period  $[t_0, t_0 + T_1(0) - 1]$ . Now we take expectations over  $T_1(0)$  on both sides conditioning on  $S(0) = s_1$  and  $\mathbf{q}(0)$ , we have:

$$\begin{aligned} \mathbb{E}\{L(T_1(0)) - L(0) \mid S(0), \mathbf{q}(0)\} \\ + \mathbb{E}\left\{\sum_{t=0}^{T_1(0)-1} V f^Q(t) \mid S(0), \mathbf{q}(0)\right\} \\ \leq C B^2 + \sum_{s_i} \mathbb{E}\{n_{s_i}^{T_1(0)}(0) \mid S(0), \mathbf{q}(0)\} g_{s_i}(\mathbf{q}(0)). \end{aligned} \quad (26)$$

Here  $C = \mathbb{E}\{\hat{C} \mid S(0), \mathbf{q}(0)\} = \frac{1}{2}[\overline{T_1^2} + \overline{T_1}]$ . The above equation uses the fact that  $g_{s_i}(\mathbf{q}(0))$  is a constant given  $\mathbf{q}(0)$ . Now by Theorem 2 in Page 154 of [11] we have that:

$$\mathbb{E}\{n_{s_i}^{T_1(0)}(0) \mid S(0), \mathbf{q}(0)\} = \frac{\pi_{s_i}}{\pi_{s_1}}. \quad (27)$$

Plug this into (26), we have:

$$\begin{aligned} \mathbb{E}\{L(T_1(0)) - L(0) \mid S(0), \mathbf{q}(0)\} \\ + \mathbb{E}\left\{\sum_{t=0}^{T_1(0)-1} V f^Q(t) \mid S(0), \mathbf{q}(0)\right\} \\ \leq C B^2 + \frac{1}{\pi_{s_1}} \sum_{s_i} \pi_{s_i} g_{s_i}(\mathbf{q}(0)). \end{aligned} \quad (28)$$

Now using (14) and (18), i.e.,  $\overline{T_1} = 1/\pi_{s_1}$  and  $g(\gamma) = \sum_{s_i} \pi_{s_i} g_{s_i}(\gamma)$ , we obtain:

$$\begin{aligned} \mathbb{E}\{L(T_1(0)) - L(0) \mid S(0), \mathbf{q}(0)\} \\ + \mathbb{E}\left\{\sum_{t=0}^{T_1(0)-1} V f^Q(t) \mid S(0), \mathbf{q}(0)\right\} \\ \leq C B^2 + \overline{T_1} g(\mathbf{q}(0)). \end{aligned} \quad (29)$$

By Corollary 1, we see that  $g(\mathbf{q}(0)) \leq V f_{av}^*$ . Thus we conclude that:

$$\begin{aligned} \mathbb{E}\{L(T_1(0)) - L(0) \mid S(0), \mathbf{q}(0)\} \\ + \mathbb{E}\left\{\sum_{t=0}^{T_1(0)-1} V f^Q(t) \mid S(0), \mathbf{q}(0)\right\} \\ \leq C B^2 + \overline{T_1} V f_{av}^*. \end{aligned} \quad (30)$$

More generally, if  $t_k = t_{k-1} + T_i(t_{k-1})$  with  $t_0 = 0$  is the

$k^{th}$  time after time 0 when  $S(t) = s_1$ , we have:

$$\begin{aligned} \mathbb{E}\{L(t_{k+1}) - L(t_k) \mid S(t_k), \mathbf{q}(t_k)\} \\ + \mathbb{E}\left\{\sum_{t=t_k}^{t_{k+1}-1} V f^Q(t) \mid S(t_k), \mathbf{q}(t_k)\right\} \\ \leq C B^2 + \overline{T_1} V f_{av}^*, \end{aligned} \quad (31)$$

Now taking expectations over  $\mathbf{q}(t_k)$  on both sides, we have:

$$\begin{aligned} \mathbb{E}\{L(t_{k+1}) - L(t_k) \mid S(t_k)\} + \mathbb{E}\left\{\sum_{t=t_k}^{t_{k+1}-1} V f^Q(t) \mid S(t_k)\right\} \\ \leq C B^2 + \overline{T_1} V f_{av}^*. \end{aligned}$$

Note that given  $S(0) = s_1$ , we have the complete information of  $S(t_k)$  for all  $k$ . Hence the above is the same as:

$$\begin{aligned} \mathbb{E}\{L(t_{k+1}) - L(t_k) \mid S(0)\} + \mathbb{E}\left\{\sum_{t=t_k}^{t_{k+1}-1} V f^Q(t) \mid S(0)\right\} \\ \leq C B^2 + \overline{T_1} V f_{av}^*. \end{aligned} \quad (32)$$

Summing the above from  $k = 0$  to  $K - 1$ , we get:

$$\begin{aligned} \mathbb{E}\{L(t_K) - L(0) \mid S(0) = s_1\} \\ + \mathbb{E}\left\{\sum_{t=0}^{t_K-1} V f^Q(t) \mid S(0) = s_1\right\} \\ \leq K C B^2 + K \overline{T_1} V f_{av}^*. \end{aligned} \quad (33)$$

Using the facts that  $|f(t)| \leq \delta_{max}$ ,  $\lceil K \overline{T_1} \rceil \leq K \overline{T_1} + 1$ ,  $L(0) = 0$  and  $L(t) \geq 0$  for all  $t$ , we have:

$$\begin{aligned} \mathbb{E}\left\{\sum_{t=0}^{\lceil K \overline{T_1} \rceil - 1} V f^Q(t) \mid S(0) = s_1\right\} \\ \leq K C B^2 + K \overline{T_1} V f_{av}^* + V \delta_{max} \\ + V \delta_{max} \mathbb{E}\{|K \overline{T_1} - t_K| \mid S(0) = s_1\}. \end{aligned} \quad (34)$$

Dividing both sides by  $V \lceil K \overline{T_1} \rceil$ , we get:

$$\begin{aligned} \frac{1}{\lceil K \overline{T_1} \rceil} \mathbb{E}\left\{\sum_{t=0}^{\lceil K \overline{T_1} \rceil - 1} f^Q(t) \mid S(0) = s_1\right\} \\ \leq \frac{C B^2 K}{V \lceil K \overline{T_1} \rceil} + f_{av}^* \frac{K \overline{T_1} V}{\lceil K \overline{T_1} \rceil} + \frac{V \delta_{max}}{\lceil K \overline{T_1} \rceil} \\ + \mathbb{E}\left\{\left|\frac{t_K - K \overline{T_1}}{K}\right| \mid S(0) = s_1\right\} \frac{K \delta_{max}}{\lceil K \overline{T_1} \rceil}. \end{aligned} \quad (35)$$

Since  $t_K = \sum_{k=0}^{K-1} T_i(t_k)$  with  $t_0 = 0$ , and each  $T_1(t_k)$  is i.i.d. distributed with mean  $\overline{T_1}$  and second moment  $\overline{T_1^2} < \infty$ , we have:

$$\begin{aligned} \left[\mathbb{E}\left\{\left|\frac{t_K - K \overline{T_1}}{K}\right| \mid S(0) = s_1\right\}\right]^2 \\ \leq \mathbb{E}\left\{\left|\frac{t_K - K \overline{T_1}}{K}\right|^2 \mid S(0) = s_1\right\} \leq \frac{\overline{T_1^2}}{K}. \end{aligned} \quad (36)$$

This implies that the term  $\mathbb{E}\left\{\left|\frac{t_K - K \overline{T_1}}{K}\right| \mid S(0) = s_1\right\} \rightarrow 0$  as  $K \rightarrow \infty$ . It is also easy to see that  $\lceil K \overline{T_1} \rceil \rightarrow \infty$  and

$\frac{K}{\lceil K\overline{T}_1 \rceil} \rightarrow \frac{1}{\overline{T}_1}$  as  $K \rightarrow \infty$ . Thus using (36) and taking a limsup as  $K \rightarrow \infty$  in (35), we have:

$$\limsup_{K \rightarrow \infty} \frac{1}{\lceil K\overline{T}_1 \rceil} \mathbb{E} \left\{ \sum_{t=0}^{\lceil K\overline{T}_1 \rceil - 1} f^Q(t) \mid S(0) = s_1 \right\} \leq \frac{CB^2}{V\overline{T}_1} + f_{av}^*.$$

Now consider the case when the starting state is  $s_j \neq s_1$ . In this case, let  $T_{j1}(0)$  be the first time the system enters state  $s_1$ . Then we see that the above argument can be repeated for the system starting at time  $T_{j1}(0)$ . The only difference is that now the “initial” backlog in this case is given by  $\mathbf{q}(T_{j1}(0))$ . Specifically, we have from (33) that:

$$\begin{aligned} & \mathbb{E}\{L(\hat{t}_K) - L(T_{j1}(0)) \mid T_{j1}(0), S(0) = s_j\} \\ & + \mathbb{E}\left\{ \sum_{t=T_{j1}(0)}^{\hat{t}_K - 1} V f^Q(t) \mid T_{j1}(0), S(0) = s_j \right\} \\ & \leq KCB^2 + K\overline{T}_1 V f_{av}^*. \end{aligned} \quad (37)$$

Here  $\hat{t}_K$  is the  $K^{th}$  return time of  $S(t)$  to  $s_1$  after time  $T_{j1}(0)$ . We thus obtain:

$$\begin{aligned} & V \mathbb{E}\left\{ \sum_{t=T_{j1}(0)}^{\hat{t}_K - 1} f^Q(t) \mid T_{j1}(0), S(0) = s_j \right\} \\ & \leq KCB^2 + K\overline{T}_1 V f_{av}^* + \mathbb{E}\{L(T_{j1}(0)) \mid T_{j1}(0), S(0) = s_j\}. \end{aligned}$$

However, since the increment of each queue is no more than  $\delta_{max}$  every time slot, we see that  $L(T_{j1}(0)) \leq [T_{j1}(0)]^2 B^2 / 2$ . Also using the fact that  $|f(t)| \leq \delta_{max}$  for all  $0 \leq t \leq T_{j1}(0)$ , we have:

$$\begin{aligned} & V \mathbb{E}\left\{ \sum_{t=0}^{\hat{t}_K - 1} f^Q(t) \mid T_{j1}(0), S(0) = s_j \right\} \\ & \leq KCB^2 + K\overline{T}_1 V f_{av}^* + [T_{j1}(0)]^2 B^2 / 2 + T_{j1}(0) V \delta_{max}. \end{aligned}$$

Now taking expectations over  $T_{j1}(0)$  on both sides, and using a similar argument as (35), we get that for every starting state  $s_j$ , we have:

$$\begin{aligned} \limsup_{K \rightarrow \infty} \frac{1}{\lceil \overline{T}_{j1} + K\overline{T}_1 \rceil} \mathbb{E}\left\{ \sum_{t=0}^{\lceil \overline{T}_{j1} + K\overline{T}_1 \rceil - 1} f^Q(t) \mid S(0) = s_j \right\} \\ \leq \frac{CB^2}{V\overline{T}_1} + f_{av}^*. \end{aligned}$$

This proves the utility part (19).

(Part B: Proof of Backlog) Now we look at the backlog performance of QLA. We similarly first assume  $S(0) = s_1$ . Recall that equation (29) says:

$$\begin{aligned} & \mathbb{E}\{L(T_1(0)) - L(0) \mid S(0), \mathbf{q}(0)\} \\ & + \mathbb{E}\left\{ \sum_{t=0}^{T_1(0)-1} V f^Q(t) \mid S(0), \mathbf{q}(0) \right\} \\ & \leq CB^2 + \overline{T}_1 g(\mathbf{q}(0)). \end{aligned} \quad (38)$$

Using the definition of  $g_c(\gamma)$  defined in (49) in Appendix A, plugging the set of  $\{\vartheta_k^{(s_i)}\}_{k=1,2,\dots,r+2}^{i=1,2,\dots}$  variables and the set of actions  $\{x_k^{(s_i)}\}_{k=1,2,\dots,r+2}^{i=1,2,\dots}$  in the slackness assumption (1),

and using the facts that  $g(\gamma) = g_c(\gamma)$  and  $0 \leq f(t) \leq \delta_{max}$ , it can be shown that  $g(\gamma)$  satisfies:

$$g(\gamma) \leq V\delta_{max} - \eta \sum_{j=1}^r \gamma_j. \quad (39)$$

Using this in (38), we have:

$$\begin{aligned} & \mathbb{E}\{L(T_1(0)) - L(0) \mid S(0), \mathbf{q}(0)\} \\ & + \mathbb{E}\left\{ \sum_{t=0}^{T_1(0)-1} V f^Q(t) \mid S(0), \mathbf{q}(0) \right\} \\ & \leq CB^2 + \overline{T}_1 V \delta_{max} - \overline{T}_1 \eta \sum_{j=1}^r q_j(0). \end{aligned} \quad (40)$$

More generally, we have:

$$\begin{aligned} & \mathbb{E}\{L(t_{k+1}) - L(t_k) \mid S(t_k), \mathbf{q}(t_k)\} \\ & \leq CB^2 + \overline{T}_1 V \delta_{max} - \overline{T}_1 \eta \sum_{j=1}^r q_j(t_k). \end{aligned} \quad (41)$$

Here  $t_k$  is the  $k^{th}$  return time of  $S(t)$  to state  $s_1$  after time 0. Taking expectations on both sides over  $\mathbf{q}(t_k)$  and rearranging the terms, we get:

$$\begin{aligned} & \mathbb{E}\{L(t_{k+1}) - L(t_k) \mid S(t_k)\} \\ & + \overline{T}_1 \eta \sum_{j=1}^r \mathbb{E}\{q_j(t_k) \mid S(t_k)\} \leq CB^2 + \overline{T}_1 V \delta_{max}. \end{aligned} \quad (42)$$

Now using the fact that conditioning on  $S(t_k)$  is the same as conditioning on  $S(0)$ , we have:

$$\begin{aligned} & \mathbb{E}\{L(t_{k+1}) - L(t_k) \mid S(0)\} \\ & + \overline{T}_1 \eta \sum_{j=1}^r \mathbb{E}\{q_j(t_k) \mid S(0)\} \leq CB^2 + \overline{T}_1 V \delta_{max}. \end{aligned} \quad (43)$$

Summing over  $k = 0, \dots, K-1$ , rearranging the terms, and using the facts that  $L(0) = 0$  and  $L(t) \geq 0$  for all  $t$ :

$$\sum_{k=0}^{K-1} \overline{T}_1 \eta \sum_{j=1}^r \mathbb{E}\{q_j(t_k) \mid S(0)\} \leq KCB^2 + K\overline{T}_1 V \delta_{max}. \quad (44)$$

Dividing both sides by  $K\overline{T}_1 \eta$ , we get:

$$\frac{1}{K} \sum_{k=0}^{K-1} \sum_{j=1}^r \mathbb{E}\{q_j(t_k) \mid S(0)\} \leq \frac{CB^2 + \overline{T}_1 V \delta_{max}}{\overline{T}_1 \eta}. \quad (45)$$

Now using the fact that  $|q_j(t+\tau) - q_j(t)| \leq \tau \delta_{max}$ , we have:

$$\begin{aligned} & \sum_{\tau=t_k}^{t_{k+1}-1} \sum_{j=1}^r q_j(\tau) \leq T_1(t_k) \sum_{j=1}^r q_j(t_k) \\ & + \left[ \frac{1}{2} (T_1(t_k))^2 - \frac{1}{2} T_1(t_k) \right] B^2. \end{aligned}$$

Taking expectations on both sides conditioning on  $S(0)$  (which is the same as conditioning on  $S(t_k)$ ), we get:

$$\begin{aligned} & \mathbb{E}\left\{\sum_{\tau=t_k}^{t_{k+1}-1} \sum_{j=1}^r q_j(\tau) \mid S(0)\right\} \\ & \leq \mathbb{E}\left\{T_1(t_k) \sum_{j=1}^r q_j(t_k) \mid S(0)\right\} + \left[\frac{1}{2}\overline{T_1^2} - \frac{1}{2}\overline{T_1}\right]B^2 \\ & = \overline{T_1}\mathbb{E}\left\{\sum_{j=1}^r q_j(t_k) \mid S(0)\right\} + \left[\frac{1}{2}\overline{T_1^2} - \frac{1}{2}\overline{T_1}\right]B^2. \end{aligned}$$

In the last step, we have used the fact that  $T_1(t_k)$  is independent of  $q(t_k)$ . Summing the above equation over  $k = 0, 1, \dots, K-1$ , we have:

$$\begin{aligned} & \mathbb{E}\left\{\sum_{t=0}^{t_K-1} \sum_{j=1}^r q_j(t) \mid S(0)\right\} \\ & \leq \sum_{k=0}^{K-1} \overline{T_1}\mathbb{E}\left\{\sum_{j=1}^r q_j(t_k) \mid S(0)\right\} + \frac{K[\overline{T_1^2} - \overline{T_1}]B^2}{2}. \end{aligned}$$

Dividing both sides by  $K$  and using (45), we have:

$$\begin{aligned} & \frac{1}{K}\mathbb{E}\left\{\sum_{t=0}^{t_K-1} \sum_{j=1}^r q_j(t) \mid S(0)\right\} \\ & \leq \frac{\overline{T_1}}{K} \sum_{k=0}^{K-1} \mathbb{E}\left\{\sum_{j=1}^r q_j(t_k) \mid S(0)\right\} + \frac{[\overline{T_1^2} - \overline{T_1}]B^2}{2} \\ & \leq \frac{CB^2 + \overline{T_1}V\delta_{max}}{\eta} + \frac{[\overline{T_1^2} - \overline{T_1}]B^2}{2}. \end{aligned}$$

Now notice that we always have  $t_K \geq K$ . Hence:

$$\begin{aligned} & \frac{1}{K} \sum_{t=0}^{K-1} \mathbb{E}\left\{\sum_{j=1}^r q_j(t) \mid S(0)\right\} \leq \frac{1}{K}\mathbb{E}\left\{\sum_{t=0}^{t_K-1} \sum_{j=1}^r q_j(t) \mid S(0)\right\} \\ & \leq \frac{CB^2 + \overline{T_1}V\delta_{max}}{\eta} + \frac{[\overline{T_1^2} - \overline{T_1}]B^2}{2}. \end{aligned}$$

This proves (20) for the case when  $S(0) = s_1$ . The case when  $S(0) = s_j \neq s_1$  can be treated in a similar way as in Part A. It can be shown that the above backlog bound still holds, as the effect of the backlog values before the first hitting time  $T_{j1}(0)$  will vanish as time increases. This proves the backlog bound (20). Theorem 2 thus follows by combining the two proofs. ■

#### APPENDIX A- PROOF OF THEOREM 1

We now prove Theorem 1. The proof idea is shown in Fig. 1, and can be described as follows: First we construct a “convexified” version of the deterministic problem (8) and show that it gives the exact value of  $Vf_{av}^*$ . We then show that the dual function  $g_c(\gamma)$  of this convexified problem, is exactly the same as the dual function  $g(\gamma)$  of (9). Hence the two dual problems have the same optimal value. We finally show that the duality gap is zero for the convexified problem by showing that its “utility-constraint” set is convex. Hence  $g(\gamma_V^*) = g_c^* = Vf_{av}^*$ , where  $g_c^*$  is the optimal value of the dual problem for the convexified problem.

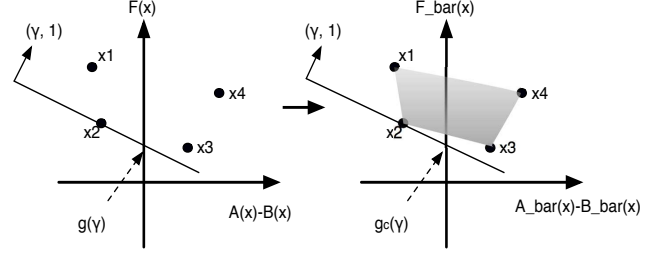


Fig. 1. The left figure shows the utility-constraint set of the deterministic problem with  $r = 1, M = 1$  and its dual function. The right figure shows the utility-constraint set of the corresponding “convexified” problem and its dual function. It can be seen that the two dual functions are the same, and that the “convexified” problem has no duality gap.

*Proof:* (Theorem 1) For notation simplicity, we denote the set of  $\mathbf{x} = (x^{(s_1)}, x^{(s_2)}, \dots), x^{(s_i)} \in \mathcal{X}^{(s_i)}$  as  $\mathcal{X}$ . We then consider the following modified deterministic problem:

$$\begin{aligned} \min : & \overline{\mathcal{F}}(\{a_k^{(s_i)}, \mathbf{x}^k\}) \triangleq V \sum_{s_i} \pi_{s_i} \sum_{k=1}^{r+2} a_k^{(s_i)} f(s_i, x_k^{(s_i)}) \quad (46) \\ \text{s.t.} & \overline{\mathcal{A}}_j(\{a_k^{(s_i)}, \mathbf{x}^k\}) \triangleq \sum_{s_i} \pi_{s_i} \sum_{k=1}^{r+2} a_k^{(s_i)} A_j(s_i, x_k^{(s_i)}) \\ & \leq \overline{\mathcal{B}}_j(\{a_k^{(s_i)}, \mathbf{x}^k\}) \triangleq \sum_{s_i} \pi_{s_i} \sum_{k=1}^{r+2} a_k^{(s_i)} \mu_j(s_i, x_k^{(s_i)}), \\ & \mathbf{x}^k \in \mathcal{X} \quad \forall k = 1, \dots, r+2, \\ & a_k^{(s_i)} \geq 0, \sum_{k=1}^{r+2} a_k^{(s_i)} = 1, \quad \forall s_i. \end{aligned}$$

Here  $\mathbf{x}^k = (x_k^{(s_1)}, x_k^{(s_2)}, \dots)$ . Due to the use of the auxiliary variables  $\{a_k^{(s_i)}\}$ , this problem can be viewed as the “convexified” version of the original deterministic problem (8). Denote the optimal value of (46) as  $OPT_c$ .<sup>2</sup> We will prove Theorem 1 via the following two claims. The first claim shows that  $OPT_c = Vf_{av}^*$  and the second claim shows that  $OPT_c = g(\gamma_V^*)$ .

*Claim 1:*  $Vf_{av}^* = OPT_c$

*Proof:* (Claim 1): For each action vector  $\mathbf{x} \in \mathcal{X}$ , we define its “utility-constraint” vector  $\mathbf{J}(\mathbf{x})$  as follows:

$$\mathbf{J}(\mathbf{x}) = (\mathcal{F}(\mathbf{x}), \mathcal{A}_1(\mathbf{x}) - \mathcal{B}_1(\mathbf{x}), \dots, \mathcal{A}_r(\mathbf{x}) - \mathcal{B}_r(\mathbf{x})).$$

Denote  $\mathcal{J} = \{\mathbf{J}(\mathbf{x}) : \mathbf{x} \in \mathcal{X}\}$ , i.e.,  $\mathcal{J}$  is the set of all possible utility-constraint vectors for  $\mathcal{X}$ , and denote  $\overline{\mathcal{J}}$  the convex hull of  $\mathcal{J}$ . Let  $\Pi^*$  be an optimal action-choosing policy that solves the stochastic problem. Now define the “utility-constraint” vector  $\mathbf{J}$  for  $\Pi^*$  as:

$$\mathbf{J}^{\Pi^*} = (\overline{f^{\Pi^*}}, \overline{A_1^{\Pi^*}} - \overline{\mu_1^{\Pi^*}}, \dots, \overline{A_r^{\Pi^*}} - \overline{\mu_r^{\Pi^*}}),$$

where  $\overline{f^{\Pi^*}} = f_{av}^*$  is the time average cost under  $\Pi^*$ , and  $\overline{A_j^{\Pi^*}}, \overline{\mu_j^{\Pi^*}}$  are the time average input and output rates to queue

<sup>2</sup>Without loss of generality, we assume such an optimal value exists. Else we can replace the “min” with “inf” in (46), consider an  $\epsilon$ -optimal solution and let  $\epsilon \rightarrow 0$ . Below we will use similar assumptions about the existence of an optimal policy for the stochastic problem, and the existence of an optimal solution to (46).

$j$  under  $\Pi^*$ . Note that here we have assumed without loss of generality that the time averages converge.<sup>3</sup> It can then be shown by using an argument similar to that in [3] that the vector  $\mathbf{J}^{\Pi^*} \in \mathcal{J}$ . Using Caratheodory's theorem [12], we see then there exist  $\{a_k^{(s_i)}\}_{i=1,2,\dots}^{k=1,\dots,r+2}$  with  $a_k^{(s_i)} \geq 0$ ,  $\sum_k a_k^{(s_i)} = 1$ , and a set of action vectors  $\{\mathbf{x}^k\}_{k=1}^{r+2} \subset \mathcal{X}$  such that:

$$\begin{aligned} \sum_{s_i} \pi_{s_i} \sum_{k=1}^{r+2} a_k^{(s_i)} f(s_i, x_k^{(s_i)}) &= \overline{f^{\Pi^*}}, \\ \sum_{s_i} \pi_{s_i} \sum_{k=1}^{r+2} a_k^{(s_i)} A_j(s_i, x_k^{(s_i)}) - \sum_{s_i} \pi_{s_i} \sum_{k=1}^{r+2} a_k^{(s_i)} \mu_j(s_i, x_k^{(s_i)}) \\ &= \overline{A_j^{\Pi^*}} - \overline{\mu_j^{\Pi^*}} \leq 0 \quad \forall j. \end{aligned}$$

The inequality  $\overline{A_j^{\Pi^*}} - \overline{\mu_j^{\Pi^*}} \leq 0$  holds since  $\Pi^*$  is by definition a stabilizing policy. This shows that  $\{a_k^{(s_i)}\}_{i=1,2,\dots}^{k=1,\dots,r+2}$ ,  $\{\mathbf{x}^k\}_{k=1}^{r+2}$  is a feasible solution of (46), implying  $V_{av}^* \geq OPT_c$ .

To prove the other direction, let  $\{a_k^{(s_i)}\}_{i=1,2,\dots}^{k=1,\dots,r+2}$  and  $\{\mathbf{x}^k\}_{k=1}^{r+2}$  be an optimal solution pair of (46). Now by our slackness assumption (1), there exists a set of actions  $\{x_k^{(s_i)}\}_{i=1,2,\dots}^{k=1,\dots,r+2}$  and probabilities  $\{\vartheta_k^{(s_i)}\}_{i=1,2,\dots}^{k=1,\dots,r+2}$  with  $\sum_k \vartheta_k^{(s_i)} = 1$  such that  $\sum_{s_i} \pi_{s_i} \{ \sum_k \vartheta_k^{(s_i)} [A_j(s_i, x_k^{(s_i)}) - \mu_j(s_i, x_k^{(s_i)})] \} \leq -\eta$  for some  $\eta > 0$  for all  $j$ . We can thus construct the following policy  $\Pi'$ : fix some  $\hat{\epsilon} \in (0, 1)$ , at every state  $s_i$ , choose action  $\tilde{x}_k^{(s_i)}$  with probability  $(1 - \hat{\epsilon})a_k^{(s_i)}$  and choose action  $x_k^{(s_i)}$  with probability  $\hat{\epsilon}\vartheta_k^{(s_i)}$ . Since  $|f(t)| \leq \delta_{max}$  for all  $t$ , it is easy to see then under  $\Pi'$ :

$$|\overline{f^{\Pi'}} - OPT_c/V| \leq \hat{\epsilon}\delta_{max}, \quad (47)$$

and that for each queue  $j$ ,  $\overline{A_j^{\Pi'}} - \overline{\mu_j^{\Pi'}} < -\hat{\epsilon}\eta < 0$ . This policy can be shown to ensure that the network is strongly stable. Hence  $\Pi'$  is a feasible control policy. Therefore  $\overline{f^{\Pi'}} \geq f_{av}^*$  by the definition of  $f_{av}^*$ . Using this fact together with (47), we have  $OPT_c/V \geq f_{av}^* - \hat{\epsilon}\delta_{max}$ . Since this holds for all  $\hat{\epsilon} > 0$ , we have  $OPT_c/V \geq f_{av}^*$ . ■

Claim 2:  $OPT_c = g(\gamma_V^*)$

*Proof:* We first look at the dual problem of (46):

$$\max : g_c(\gamma), \quad s.t. \quad \gamma \succeq \mathbf{0}, \quad (48)$$

where  $g_c(\gamma)$  is defined:

$$\begin{aligned} g_c(\gamma) = & \inf_{x_k^{(s_i)} \in \mathcal{X}^{(s_i)}, a_k^{(s_i)}} \sum_{s_i} \pi_{s_i} \left\{ V \sum_{k=1}^{r+2} a_k^{(s_i)} f(s_i, x_k^{(s_i)}) \right. \\ & \left. + \sum_j \gamma_j \left[ \sum_{k=1}^{r+2} a_k^{(s_i)} A_j(s_i, x_k^{(s_i)}) - \sum_{k=1}^{r+2} a_k^{(s_i)} \mu_j(s_i, x_k^{(s_i)}) \right] \right\}. \end{aligned} \quad (49)$$

Now by comparing (49) and (10), we see that  $g_c(\gamma) = g(\gamma)$  for all  $\gamma \succeq \mathbf{0}$ . This is so because at any  $\gamma \succeq \mathbf{0}$ , we first have  $g_c(\gamma) \leq g(\gamma)$ . Now if  $\{x_k^{(s_i)}\}_{i=1,2,\dots}^\infty$  are the minimizers of  $g(\gamma)$ , then  $\{x_k^{(s_i)}\}_{i=1,2,\dots}^{k=1,\dots,r+2}$ , with  $x_k^{(s_i)} = x_k^{(s_i)}$ ,  $a_1^{(s_i)} = 1$  and  $a_k^{(s_i)} = 0$  if  $k \neq 1$ , will also be the minimizers of  $g_c(\gamma)$ . This shows  $g_c(\gamma) = g(\gamma)$ , which then implies that  $g_c^* = g(\gamma_V^*)$ , where  $g_c^*$  is the optimal value of (48).

<sup>3</sup> In the case when this assumption is violated, the same argument can be applied to the limit points of the time averages but is more involved.

Now it remains to show that  $g_c^* = OPT_c$ . It suffices to show that  $g_c^* \geq OPT_c$ . We prove this claim by using a similar approach as that in Page 234 of [15]. Denote the set  $\Gamma = \{\{a_k^{(s_i)}\}_{i=1,2,\dots}^{k=1,\dots,r+2} : a_k^{(s_i)} \geq 0, \sum_k a_k^{(s_i)} = 1, \forall s_i\}$ . Consider the set  $\mathcal{M}$  as follows:

$$\begin{aligned} \mathcal{M} = & \left\{ (u, c_1, \dots, c_r) \mid \exists \{a_k^{(s_i)}\} \in \Gamma, \{\mathbf{x}^k\}_{k=1}^{r+2} \subset \mathcal{X} \text{ s.t.} \right. \\ & \overline{\mathcal{F}}(\{a_k^{(s_i)}, \mathbf{x}^k\}) \leq u, \text{ and} \\ & \left. \overline{\mathcal{A}}_j(\{a_k^{(s_i)}, \mathbf{x}^k\}) - \overline{\mathcal{B}}_j(\{a_k^{(s_i)}, \mathbf{x}^k\}) \leq c_j, \forall j \right\}. \end{aligned}$$

It is not difficult to show that  $\overline{\mathcal{F}} \subset \mathcal{M}$ . We now show that  $\mathcal{M}$  is convex. Indeed, if two vectors  $(u, c_1, \dots, c_r)$  and  $(\hat{u}, \hat{c}_1, \dots, \hat{c}_r)$  are both in  $\mathcal{M}$ , then there exist  $\{a_k^{(s_i)}\}_{i=1,2,\dots}^{k=1,\dots,r+2}$ ,  $\{\mathbf{x}^k\}_{k=1}^{r+2}$  and  $\{\hat{a}_k^{(s_i)}\}_{i=1,2,\dots}^{k=1,\dots,r+2}$ ,  $\{\hat{\mathbf{x}}^k\}_{k=1}^{r+2}$  such that:

$$\begin{aligned} \overline{\mathcal{F}}(\{a_k^{(s_i)}, \mathbf{x}^k\}) &\leq u, \quad \overline{\mathcal{F}}(\{\hat{a}_k^{(s_i)}, \hat{\mathbf{x}}^k\}) \leq \hat{u}, \\ \overline{\mathcal{A}}_j(\{a_k^{(s_i)}, \mathbf{x}^k\}) - \overline{\mathcal{B}}_j(\{a_k^{(s_i)}, \mathbf{x}^k\}) &\leq c_j, \forall j, \\ \overline{\mathcal{A}}_j(\{\hat{a}_k^{(s_i)}, \hat{\mathbf{x}}^k\}) - \overline{\mathcal{B}}_j(\{\hat{a}_k^{(s_i)}, \hat{\mathbf{x}}^k\}) &\leq \hat{c}_j, \forall j. \end{aligned}$$

Now if we consider the vectors  $\theta \cdot (u, c_1, \dots, c_r) + (1 - \theta) \cdot (\hat{u}, \hat{c}_1, \dots, \hat{c}_r)$ . Using Caratheodory's theorem again, we see that there exists  $\{\tilde{a}_k^{(s_i)}\}_{i=1,2,\dots}^{k=1,\dots,r+2}$ ,  $\{\tilde{\mathbf{x}}^k\}_{k=1}^{r+2}$  such that:

$$\begin{aligned} \overline{\mathcal{F}}(\{\tilde{a}_k^{(s_i)}, \tilde{\mathbf{x}}^k\}) &= \theta \overline{\mathcal{F}}(\{a_k^{(s_i)}, \mathbf{x}^k\}) + (1 - \theta) \overline{\mathcal{F}}(\{\hat{a}_k^{(s_i)}, \hat{\mathbf{x}}^k\}), \\ \overline{\mathcal{A}}_j(\{\tilde{a}_k^{(s_i)}, \tilde{\mathbf{x}}^k\}) - \overline{\mathcal{B}}_j(\{\tilde{a}_k^{(s_i)}, \tilde{\mathbf{x}}^k\}) &= \theta \left[ \overline{\mathcal{A}}_j(\{a_k^{(s_i)}, \mathbf{x}^k\}) - \overline{\mathcal{B}}_j(\{a_k^{(s_i)}, \mathbf{x}^k\}) \right] \\ &\quad + (1 - \theta) \left[ \overline{\mathcal{A}}_j(\{\hat{a}_k^{(s_i)}, \hat{\mathbf{x}}^k\}) - \overline{\mathcal{B}}_j(\{\hat{a}_k^{(s_i)}, \hat{\mathbf{x}}^k\}) \right]. \end{aligned}$$

This implies that  $\theta \cdot (u, c_1, \dots, c_r) + (1 - \theta) \cdot (\hat{u}, \hat{c}_1, \dots, \hat{c}_r) \in \mathcal{M}$ , hence  $\mathcal{M}$  is convex.

We now define a second convex set  $\mathcal{D}$  as  $\mathcal{D} = \{(\nu, c_1, \dots, c_r) \mid \nu < OPT_c, c_j = 0, \forall j\}$ . It is easy to see then  $\mathcal{M} \cap \mathcal{D}$  is empty, for otherwise  $OPT_c$  can not be the optimal value of (48). Therefore there exists a hyperplane with norm  $(\zeta, \gamma_1, \dots, \gamma_r) \neq \mathbf{0}$  and some constant  $c$  such that:

$$\begin{aligned} (u, c_1, \dots, c_r) \in \mathcal{M} &\Rightarrow u\zeta + \sum_{k=1}^r \gamma_k c_k \geq c, \\ (\nu, c_1, \dots, c_r) \in \mathcal{D} &\Rightarrow \nu\zeta + \sum_{k=1}^r \gamma_k c_k \leq c. \end{aligned} \quad (50)$$

We can thus conclude that  $\zeta \geq 0, \gamma_j \geq 0$  and  $\zeta\nu \leq c$  for all  $\nu < OPT_c$ , which implies  $\zeta OPT_c \leq c$ . Using these in (50), and using the fact that  $\overline{\mathcal{F}} \subset \mathcal{M}$ , we see that for any  $\{a_k^{(s_i)}\}_{i=1,2,\dots}^{k=1,\dots,r+2} \in \Gamma$ ,  $\{\mathbf{x}^k\}_{k=1}^{r+2} \subset \mathcal{X}$ , we have:

$$\begin{aligned} \zeta OPT_c \leq c &\leq \zeta \overline{\mathcal{F}}(\{a_k^{(s_i)}, \mathbf{x}^k\}) \\ &\quad + \sum_{j=1}^r \gamma_j \left[ \overline{\mathcal{A}}_j(\{a_k^{(s_i)}, \mathbf{x}^k\}) - \overline{\mathcal{B}}_j(\{a_k^{(s_i)}, \mathbf{x}^k\}) \right]. \end{aligned} \quad (51)$$

Clearly,  $\zeta \neq 0$ , for otherwise we can plug in the actions  $\{x_k^{(s_i)}\}_{i=1,2,\dots}^{k=1,\dots,r+2}$  and probabilities  $\{\vartheta_k^{(s_i)}\}_{i=1,2,\dots}^{k=1,\dots,r+2}$  in the



slackness assumption (1) to obtain:

$$0 \geq \sum_{j=1}^r \gamma_j (-\eta) \geq 0,$$

which will imply  $(\zeta, \gamma_1, \dots, \gamma_r) = \mathbf{0}$ . Thus we see that  $\zeta > 0$ . Now dividing  $\zeta$  from both sides of (51), we have:

$$\begin{aligned} \overline{\mathcal{F}}(\{a_k^{(s_i)}, \mathbf{x}^k\}) + \sum_{j=1}^r \gamma'_j \left[ \overline{\mathcal{A}}_j(\{a_k^{(s_i)}, \mathbf{x}^k\}) - \overline{\mathcal{B}}_j(\{a_k^{(s_i)}, \mathbf{x}^k\}) \right] \\ \geq OPT_c, \end{aligned}$$

where  $\gamma'_j = \gamma_j / \zeta$ . This implies  $g_c(\gamma') \geq OPT_c$  with  $\gamma' = (\gamma'_1, \dots, \gamma'_r)^T$ . Hence  $g_c^* \geq g_c(\gamma') \geq OPT_c$ , which by weak duality implies  $g_c^* = OPT_c$ , and so  $g(\gamma_V^*) = OPT_c$ . ■

Combining Claim 1 and 2, we see that Theorem 1 follows. ■

## REFERENCES

- [1] L. Georgiadis, M. J. Neely, and L. Tassiulas. *Resource Allocation and Cross-Layer Control in Wireless Networks*. Foundations and Trends in Networking Vol. 1, no. 1, pp. 1-144, 2006.
- [2] A. Eryilmaz and R. Srikant. Fair resource allocation in wireless networks using queue-length-based scheduling and congestion control. *IEEE/ACM Trans. Netw.*, 15(6):1333–1344, 2007.
- [3] M. J. Neely. Energy optimal control for time-varying wireless networks. *IEEE Transactions on Information Theory* 52(7): 2915-2934, July 2006.
- [4] L. Huang and M. J. Neely. The optimality of two prices: Maximizing revenue in a stochastic network. *Proc. of 45th Annual Allerton Conference on Communication, Control, and Computing (invited paper)*, Sept. 2007.
- [5] R. Urgaonkar and M. J. Neely. Opportunistic scheduling with reliability guarantees in cognitive radio networks. *IEEE INFOCOM Proceedings*, April 2008.
- [6] Y. Yi and M. Chiang. Stochastic network utility maximization: A tribute to kelly's paper published in this journal a decade ago. *European Transactions on Telecommunications*, vol. 19, no. 4, pp. 421-442, June 2008.
- [7] M. J. Neely. Super-fast delay tradeoffs for utility optimal fair scheduling in wireless networks. *IEEE Journal on Selected Areas in Communications (JSAC), Special Issue on Nonlinear Optimization of Communication Systems*, 24(8), Aug. 2006.
- [8] M. J. Neely. Optimal energy and delay tradeoffs for multi-user wireless downlinks. *IEEE Transactions on Information Theory* vol. 53, no. 9, pp. 3095-3113, Sept. 2007.
- [9] L. Huang and M. J. Neely. Delay reduction via lagrange multipliers in stochastic network optimization. *Proc. of WiOpt, Seoul*, June 2009.
- [10] M. J. Neely. Stability and capacity regions of discrete time queueing networks. *arXiv:1003.3396v1*, March 2010.
- [11] R. G. Gallager. *Discrete Stochastic Processes*. Kluwer Academic Publishers, 1996.
- [12] D. P. Bertsekas, A. Nedic, and A. E. Ozdaglar. *Convex Analysis and Optimization*. Boston: Athena Scientific, 2003.
- [13] M. J. Neely, E. Modiano, and C. E. Rohrs. Dynamic power allocation and routing for time-varying wireless networks. *IEEE Journal on Selected Areas in Communications*, Vol 23, NO.1, January 2005.
- [14] M. J. Neely. Delay analysis for maximal scheduling with flow control in wireless networks with bursty traffic. *IEEE Transactions on Networking*, August 2009.
- [15] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.